An Overview of Generative Adversarial Networks

Xinyue Long, Mingchuan Zhang

School of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

Abstract: Generative adversarial network (GAN), put forward by two-person zero-sum game theory, is one of the most important research hotspots in the field of artificial intelligence. With a generator network and a discriminator network, GAN is trained by adversarial learning. In this paper, we aim to discusses the development status of GAN. We first introduce the basic idea and training process of GAN in detail, and summarize the structure and structure of GAN derivative models, including conditional GAN, deep convolution DCGAN, WGAN based on Wasserstein distance and WGAN-GP based on gradient strategy. We also introduce the specific applications of GAN in the fields of information security, face recognition, 3D and video technology, and summarize the shortcomings of GAN. Finally, we look forward to the development trend of GAN.

Keywords: Generative Adversarial Network; Derivative Models; Development Trend.

1. Introduction

In recent years, the new generation of artificial intelligence technology, represented by deep learning, has developed rapidly. Deep learning stands out from traditional machine learning due to its strong ability of data reduction and analysis, nonlinear fitting and feature extraction. It can extract key information from original data samples containing several complex characteristics by directly learning the probability distribution of data samples, and is not limited by the strength of the feature expression ability of the data itself [1]. At present, deep learning models are divided into two categories: generation model and discrimination model. The discrimination model develops rapidly with the invention of algorithms such as backpropagation and dropout, but the generation model has no major breakthrough due to the difficulty of data modeling [2]. It was not until Goodfellow et al. [3] proposed a generative adversarial network based on the zero-sum game idea, which attracted attention from academia and industry as soon as it was proposed. The relevant theories of GAN have been widely applied in the fields of image generation [4], image restoration [5], image recognition [6], natural language processing [7], etc., and are constantly extended to other fields.

The structure that generated the adversarial network was inspired by the two-person zero-sum game problem in game theory: the sum of two people's interests is zero, and one person's gain is the other's loss. The network consists of generator model and discriminator model. The generated model is used to obtain the real data distribution and judge whether the model prediction input comes from the real data or the data generated by the generated model. The two are trained alternately in the network, learn and progress together in the confrontation, and finally reach the Nash equilibrium state. Its network model structure is shown in Figure 1.
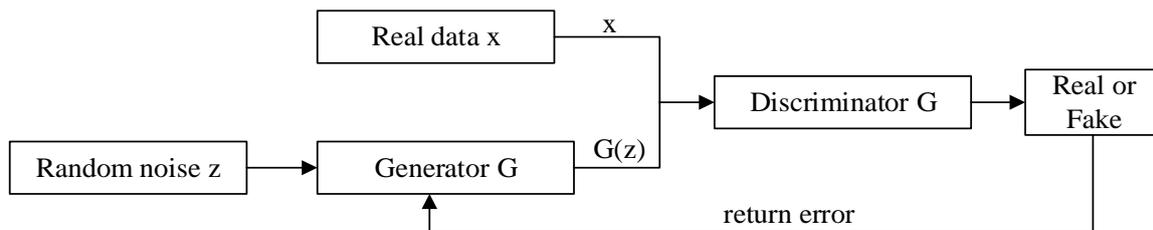

Fig.1 Basic architecture diagram of GAN

GAN inputs real data x and random variable z, and trains both models G and D at the same time. The purpose of G is to make $D(G(z))$ large enough to maximize the possibility of correct labels from generated samples, so as to generate the same data as real data as much as possible. The discriminator cannot judge that the generated data $G(z)$ is false data; The purpose of D is to make $D(G(z))$ small enough to correctly judge whether the input data is real data or fake data as much as possible. The discriminator inputs a number between 0 and 1, which represents the probability that the input data is from real data. The input types of the discriminator are (x_{real}, 1) and (x_{fake}, 0). After several confrontation adjustments, the generator and discriminator finally reach a dynamic Nash equilibrium.

The specific training process is as follows: the generator inputs random noise, the discriminator inputs real data and the data generated by the generator. During the training process, one side of the generator or the discriminator is fixed and the weight of the other side is updated. The discriminator updates with D data of formula (1) random gradient ascending, while the generator G data of formula (2) random gradient descending.

where x represents real data, z is the random variable, $G(z)$ represents the samples generated by the generator that obey the real data distribution, $D(x)$ represents the probability of judging that the input data is from real data, $D(G(z))$ represents the probability that the input data is from the generated fake data. Generator and discriminator are iterated alternately, in which G and D try their best to optimize their networks, forming a state of competition and

confrontation until the model converges and both G and D reach Nash equilibrium. However, at the early stage of training, if the generation effect of the generator is poor, the discriminator will refuse to generate samples with high confidence. $log(1 - D((z)))$ can easily reach saturation, so choose to maximize $logD(G(z))$ rather than minimize $log(1 - D((z))$. The crossentropy loss function is used to calculate the loss of the discriminator, the loss function of the discriminator:

the loss function of the generator:

the final loss function:

After updating the discriminator parameters, it update the generator parameters. The competitive adversarial optimization process of GAN can be described as:

$$\nabla\theta d \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \right] \tag{1}$$

$$\nabla\theta g \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{i}\right)\right)\right) \tag{2}$$

$$-((1-y)\log(1 - D(G(z))) + y\log D(x)) \tag{3}$$

$$(1-y)\log(1 - D(G(z))) \tag{4}$$

$$(1-y)\log(1 - D(G(z)))(2 * \bar{D}(G(z)) - 1) \tag{5}$$

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{6}$$

where $x \sim p_{data}(x)$ indicates that x follows the real data distribution, and $z \sim p(z)$ indicates that x follows the real data distribution z obeys Gaussian distribution, and $\mathbb{E}(*)$ is the data expectation. When $p_g = p_{data}$, the network model reaches the best result. At this time, the generator generates the same sample as the real data, and the discriminator can no longer distinguish whether the data is the real data or the data generated by the generator.

2. Derivative model of GAN

Generating adversarial network model GAN, as a popular network model, has been discussed a lot. It has the following advantages: this model is combined with two neural networks for adversarial training, and only uses back propagation instead of complex Markov chain [8], which improves the training efficiency; there is no need to infer hidden variables during training, which reduces the difficulty of training. Although GAN has many advantages, it also has disadvantages that cannot be ignored: it is found that GAN is not suitable for processing discrete forms of data; Gradient descent [9] is sometimes effective, but sometimes difficult to achieve ideal results; The balance and synchronization of the two antagonistic networks should be ensured in the training process, which is too difficult and easy to cause instability in the training process. Therefore, because the traditional generative adversarial network model is relatively simple and has poor generative effect, a large number of studies have improved it, mainly in the optimization of network architecture, loss function and training model.

CGAN [10], as a supervised model, adds additional one-hot vector c on the input layer of generator and discriminator as the iterative optimization direction of the network constrained by conditional information to realize directional data generation, that is, given conditional constraints to make the network output conform to the preset results. CGAN and its derived model based on conditional constraints alleviate the phenomenon that GAN networks generate data too freely by directively generating data, and help eliminate the phenomenon of schema collapse, its network model structure is shown in Figure 2. For example, in the scenario where the power system meets the prediction, the condition information can be the factor affecting the load. During the generation of renewable energy scene, the conditional information can be the spatio-temporal information of renewable energy output, such as date, location coordinates, etc. For power equipment fault diagnosis, the condition information can be specific fault category.
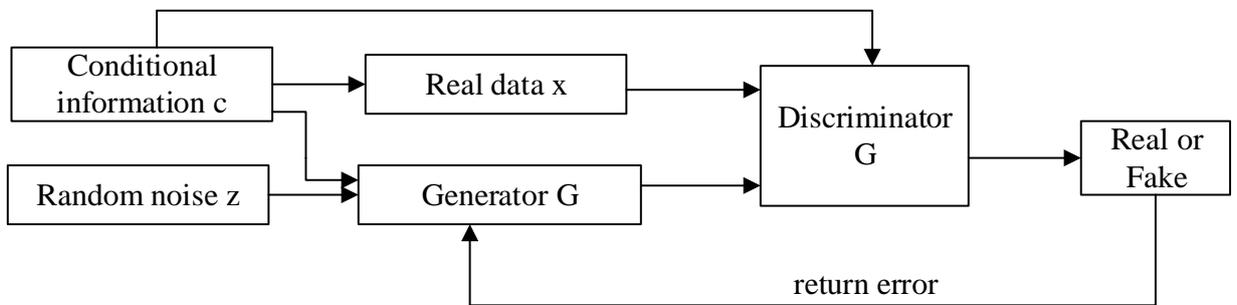

Fig.2 Basic architecture diagram of CGAN

DCGAN [11], modified its network structure by referring to the principle of the original GAN. Convolutional neural network (CNN) became the structural subject of generator G and discriminator D, and the quality and convergence rate of generated samples were greatly improved, its network model structure is shown in Figure 3. In G network and D network, micro-step amplitude convolution and step length convolution are respectively used to replace the previous pooling layer. Different from the common CNN used to extract features, CNN structure in DCGAN needs to generate

images, and pooling will ignore a lot of information. For the problem of gradient disappearance, batch standardization can be used in D and G; The use of full connection layer is abandoned and the network becomes full convolutional network. LeakyReLU is used as the activation function in network G, tanh is the activation function at the last layer, and Leakyrelu is used as the activation function in network D. DCGAN is widely used in image generation, which improves the training stability of GAN and the quality of generated results to a large extent. However, it only improves the structure of GAN, but does not fundamentally solve the problem of training stability of GAN. In the training process, it is still necessary to balance the training times of generator and discriminator.

WGAN, was proposed by Arjovsky et al. [12] for the problem of divergence distance measurement. The Earth Mover distance (EM distance) was used to calculate the distance between the two distributions, and the EM distance was used to monitor the quality of the model, so as to solve the problems of GAN training instability, model collapse and evaluation of the generated model, the structure of WGAN is shown in Figure 4. When dealing with Lipschitz constraints, WGAN adopted the weight clipping method to limit the gradient value of parameters within a certain range, so as to limit the growth rate of D(x) by reciprocal. However, the weight clipping value in this method was not easy to determine, which sometimes only generated poor samples or failed to converge. In fact, WGAN cannot completely solve the problem of GAN training stability. The Lipschitz restriction method in WGAN needs to truncate the absolute value of the discriminator parameter to no more than a fixed constant c.
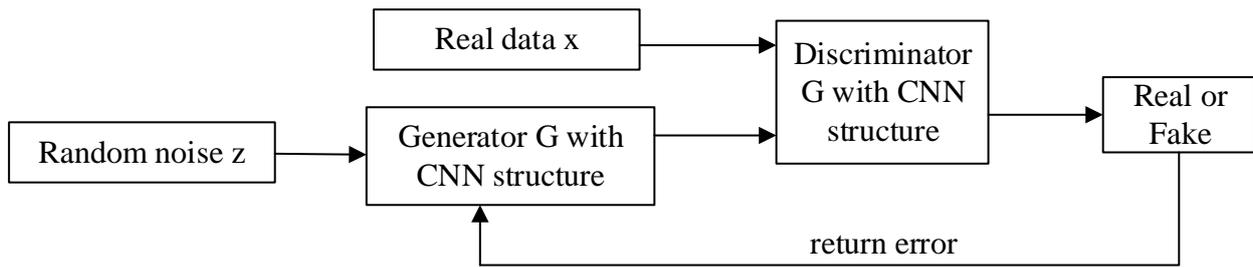

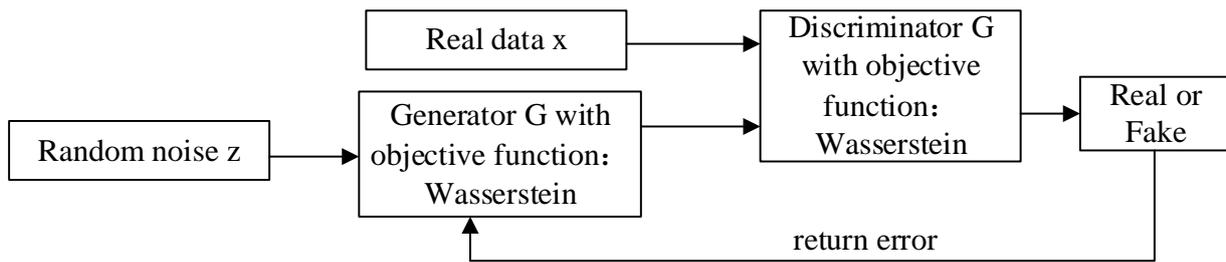**Fig.3** Basic architecture diagram of DCGAN

Fig.4 Basic architecture diagram of WCGAN

In view of the shortcomings of WGAN, Gulrajani et al [13] proposed an improved version of WGAN, named WGAN-GP. WGAN-gp proposed a new Lipschitz continuity restriction technique, that is, by adding a gradient penalty term to the discriminant function, the parameters are associated with the restriction. The Lipschitz restriction condition is reached, so as to solve the problems of parameter centralization, gradient explosion and gradient disappearance caused by weight cutting of WGAN. However, although Wgan-GP also penalizes the values in areas where the gradient magnitude is greater than 1, it cannot guarantee that the gradient magnitude of each value is less than or equal to 1.

3. Application of GAN

3.1. Information security applications

As the application of GANs continues to expand, it also has some applications in privacy protection. In data privacy protection, how to ensure the balance between data set availability and privacy is very important. GANs takes advantage of its ability to add noise to the underlying space rather than directly to the data, reducing overall information loss while ensuring privacy. Triastcyn et al. [14] proposed a method to generate artificial data sets. By adding Gaussian noise layer to the discriminator of generating adversarial networks, the output and gradient have different privacy compared with the training data. It also provides differential privacy protection for these data. Huang et al. [15] proposed a context-aware privacy model in combination with GANs, which realized the release of private data by tactfully adding noise. Frigerio et al. [16] proposed a privacy protection data release framework through the definition of differential privacy. From time series to generation of continuous data and discrete data, it can be easily adapted to different use cases, so as to ensure the release of new open data while protecting the user's personality.

In addition to privacy protection, GANs is also used in malicious detection. In order to effectively detect malware including zero-day attacks, Kim et al. [17] proposed a transferred deep-convolutional generation adversarial network (tDCGAN), which is based on deep self-coding technology, the actual data and modified data generated by tDCGAN are used to learn the features of various malware, and meaningful features are extracted for malware detection. GANs is also applicable to fraud detection of credit cards. Fiore et al. [18] trained a GANs model to output simulated cheating examples of a few classes, and then combined these examples with training data into an enhanced training set, so as to improve the classification efficiency of the classifier for cheating examples of a few classes. In addition to these malicious detections, GANs can also be used to detect botnets. Yin et al. [19] proposed a botnet detection model framework

based on GANs -- Bot-GAN. The generated model constantly generates pseudo-samples to assist the original detection model to improve its performance.

The wide applicability of GANs makes it applicable to the oldest cryptography in the field of information security. Abadi et al. [20] used the adversarial learning mechanism of GANs to replace the communication parties and adversaries in the traditional symmetric encryption system with neural networks, so as to realize the process of encryption and decryption and achieve the purpose of protecting the communication process. Gomez et al. [21] proposed CipherGAN based on GANs to decipher shift ciphergan and Virginian ciphergan in classical cryptography. Hitaj et al. [22] proposed PassGAN, a new method to enhance cryptography with GANs. By training GANs in the list of leaked ciphers to crack ciphers, GANs can be applied in cryptography in a broader way.

3.2. Face recognition

Last year, Ukrainian special forces made an ambush plan and used facial recognition technology to locate Russian Lieutenant General Andrei Sicheva by mastering his face, pupil, voice print and other biometric features. In this special operation, face recognition technology plays an important role. The deep application of face recognition technology in military field has attracted wide attention from other countries and military organizations. Zeng Fanzhi et al. [23] proposed a GAN method to improve face image clarity. In this model, global depth convolution is introduced to solve the problem that the weight of convolution kernel in different regions is consistent in face image, which improves the accuracy of face recognition. Marriott et al. [24] proposed to combine 3D variable models into GAN generators, and control the pose, lighting and expression of images if conditions permit. 3D variability GAN can accurately reconstruct the 3D structure of a human face. The introduction of GAN greatly improves the accuracy and accuracy of face recognition in computer vision. For face recognition task, the change of people's age will have a great impact on its judgment accuracy, the reason is that as people grow older, facial aging will cause the change of facial features.

Therefore, face recognition systems need aging faces to improve the efficiency of feature extraction. Traditional face aging methods require researchers to understand the biological knowledge of facial aging, which requires high prior experience. The face aging method based on GAN can learn features from a large number of aging face images and generate aging face images simply and efficiently. Antipov et al. [25] proposed an age-Cgan method to generate face images under different Age conditions and retain the identity features used in training. In order to keep the identity information contained in the image, Age-CGAN uses a coder to extract the identity features to ensure that the reconstructed image can maintain the original identity information. GAN images generated by random noise often have problems such as gradient disappearance and training instability. The introduction of automatic encoder (AE) [26] has become a way to solve these problems. Zhang et al. [27] proposed a method based on conditional adversarial autoencoder (CAAE) to model faces on age-related high-dimensional fluids, achieving smooth aging and regression on face images.

3.3. 3D field

The main application scenario of GAN in the 3D field is object reconstruction. The method of 3D reconstruction is mainly through the reconstruction of the voxel of two-dimensional image surface. In 2016, Wu et al. [28] took the lead in using GAN model to solve the problem of object reconstruction in the 3D field and called it 3D-GAN. They used 3D convolution and reconstructed 3D objects from probability space according to the surface voxels of images. Henzler et al. [29] proposed PlatonicGAN, also a voxel-based approach to 3D object generation through a series of effectively separable rendering layers. In the same year, Nguyen-Phuoc et al. [30] proposed HoLoGAN, which was able to learn three-dimensional representation of images without supervision by learning rigid-body transformation of three-dimensional features. Although the above two methods can generate relatively good quality images, the PlatonicGAN method will limit the resolution of the generated images, and artifacts will be generated when the resolution increases. HoLoGAN needs to learn extra how to render images.

Recently, Mildenhall et al. [31] proposed the neural radiation field (Nerf) model, which can use neural networks to represent static objects and render clear images from any Angle after the completion of network training. Based on this method, Schwarz et al. [32] proposed the generated radiant field (Graf) method, which improved the shortcoming of Nerf requiring a large number of pictures taken from different angles of the same object with camera position and attitude. However, this method could not deal with multiple objects in the scene and had certain limitations. Niemeyer et al. [33] improved the Graf method and put forward the Giraffe method. This method combined the generation of radiation fields to generate as many radiation fields as the number of objects in a scene, and finally combined all feature images and rendered the colors of the images. Up to now, the combination of 3D and GAN is still not sufficient, and there are still challenges in the future research work in this area.

3.4. Video field

GAN can generate high-quality and realistic images, and video is composed of continuous multi-frame images. Therefore, more and more scholars try to combine generative adversarial network with video generation. VideoGAN Videogan video generation is mainly composed of continuous frame by frame image sequence. Vondrick et al. [34] propose videogan to videogan generation in the field of computer vision, which focuses on predicting the next frame of video over time. Later, Zhou et al. [35] improved the predicted timing sequence by using the timing modeling capability of RNN network on this basis, and proposed the framework of RNn-Gan to input one frame of image and predict the video of the next frame instead of random noise. Xiong et al. [36] proposed that the first stage mainly focuses on the authenticity of video frames, while the second stage mainly reflects the correlation between frames. The structure of the first stage uses the strategy of Unet structure jump link, the second stage uses the Gram matrix to maintain the movement of objects between frames, MDGAN model not only improves the clarity of the video, but also improves the dynamic sense of the picture.

Clark et al. [37] proposed Dual video discriminator GAN (DVD-GAN), which can generate longer and higher quality videos and prove that more complex and fidelity videos can be generated than previous methods on complex data sets.

The task of generating high resolution video frequencies has progressed rapidly, but achieving comparable video generation remains a major problem, Tian et al. [38] proposed MocoGAN-HD (Motion and content GAN for high definition) to conduct cross-domain video synthesis by introducing a motion generator. In addition, Skorokhodov et al. [39] proposed on the basis of StyleGAN v2 structure that Video with StyleGAN (Video with styleGAN) improved the problem that the previous work failed to stabilize the generation rate, and could fix the generation rate of arbitrarily long high-resolution videos at 64 frames.

4. The development trend of GAN

Although GANs has been applied in many aspects, it still has a broad application prospect. In the future work, the development trend of GANs can be further studied from the following points:

In the field of information security: using the characteristics of GANs, combined with image watermarking technology, the information is attached to the image for transmission, to achieve the effect of safe message transmission; By using GANs's countermeasure learning mechanism, virus samples were analyzed and new viruses were predicted to form an automatic defense system. By combining GANs with classical cryptography, the decoding rate is analyzed, and the adversarial-learning mechanism of GANs is used to ensure the secure communication under the public key cryptosystem, so as to realize the authenticability and confidentiality of messages.

Unsupervised nature: In the field of image conversion, it can be divided into supervised network and unsupervised network according to whether pairing data set training is required. Compared with supervised GAN, unsupervised GAN can handle more types of tasks due to its wide application, strong generalization and no need for paired data set training. However, from the perspective of image quality generation, supervised GAN is still better than unsupervised GAN. Therefore, under the premise of maintaining the unsupervised nature of the network, optimizing the quality of image generation is the mainstream development route of image conversion networks in the future.

Acknowledgements

References

[1] S. Zhenguo, Z. Chengsheng, and C. Feixiong, "Review of generative adversarial networks and their applications in power systems," Proceedings of the CSEE, vol. 43, no. 03, pp. 987–1004, 2023.

[2] Z. Enqi, G. Guanghua, and Z. Chen, "Research progress of generative adversarial network gan," Application Research of Computers, vol. 38, no. 04, pp. 968–974, 2021.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in Neural Information Processing Systems, Curran Associates, Inc., vol. 27, 2014.

[4] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," CoRR, vol. abs/1710.10196, 2017.

[5] D. Pathak, P. Kr¨ahenb¨uhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, pp. 2536–2544.

[6] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, pp. 5967–5976.

[7] K. Lin, D. Li, X. He, M. Sun, and Z. Zhang, "Adversarial ranking for language generation," in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, pp. 3155–3165.

[8] R. M. Neal, "Markov chain sampling methods for dirichlet process mixture models," Journal of Computational and Graphical Statistics, vol. 9, no. 02, pp. 229–265, 2012.

[9] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," IEEE Trans.Image Process., vol. 20, no. 7, pp. 2030–2048, 2011.

[10] M. Mirza and S. Osindero, "Conditional generative adversarial nets," CoRR, vol. abs/1411.1784, 2014.

[11] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings, Y. Bengio and Y. LeCun, Eds., 2016.

[12] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," CoRR, vol. abs/1701.07875, 2017.

[13] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds., 2017, pp. 5767–5777.

[14] A. Triastcyn and B. Faltings, "Generating differentially private datasets using gans," CoRR, vol. abs/1803.03148, 2018.

[15] C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Rajagopal, "Generative adversarial privacy," CoRR, vol. abs/1807.05306, 2018.

[16] L. Frigerio, A. S. de Oliveira, L. Gomez, and P. Duverger, "Differentially private generative adversarial networks for time series, continuous, and discrete open data," CoRR, vol. abs/1901.02477, 2019.

[17] J. Kim, S. Bu, and S. Cho, "Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders," Inf. Sci., vol.460-461, pp. 83–102, 2018.

[18] U. Fiore, A. D. Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using generative adversarial networks for improving classification effectiveness in creditcard fraud detection," Inf. Sci., vol. 479, pp. 448–455, 2019.

[19] C. Yin, Y. Zhu, S. Liu, J. Fei, and H. Zhang, "An enhancing framework for botnet detection using generative adversarial networks," in 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), 2018, pp. 228–234.

[20] M. Abadi and D. G. Andersen, "Learning to protect communications with adversarial neural cryptography," CoRR, vol. abs/1610.06918, 2016.

[21] A. N. Gomez, S. Huang, I. Zhang, B. M. Li, M. Osama, and L. Kaiser, "Unsupervised cipher cracking using discrete gans," in 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. OpenReview.net, 2018.

[22] B. Hitaj, P. Gasti, G. Ateniese, and F. Pérez-Cruz, "Passgan: A deep learning approach for password guessing," CoRR, vol. abs/1709.00440, 2017.

[23] Z. Y. e. a. Zeng Fanzhi, Zou Lei, "Application of conditional gan deblurring algorithm in face recognition," Small microcomputer system, vol. 42, no. 12, pp. 2607–2613, 2021.

[24] R. T. Marriott, S. Romdhani, and L. Chen, "A 3d GAN for improved large-pose facial recognition," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021. pp.13445–13455.

[25] G. Antipov, M. Baccouche, andJ. Dugelay, "Face aging with conditional generative adversarial networks," in 2017 IEEE International Conference on Image Processing, ICIP 2017, Beijing, China, September 17-20, 2017. IEEE, 2017, pp. 2089–2093.

[26] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," in Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016, ser. JMLR Workshop and Conference Proceedings, M. Balcan and K. Q. Weinberger, Eds., vol. 48. JMLR.org, 2016, pp. 1558–1566.

[27] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," CoRR, vol. abs/1702.08423, 2017.

[28] J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum, "Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling," in Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain, D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, Eds., 2016, pp. 82–90.

[29] P. Henzler, N. J. Mitra, and T. Ritschel, "Escaping plato's cave: 3d shape from adversarial rendering," in 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019.IEEE, 2019, pp. 9983–9992.

[30] T. Nguyen-Phuoc, C. Li, L. Theis, C. Richardt, and Y. Yang, "Hologan: Unsupervised learning of 3d representations from natural images," in 2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, Seoul, Korea (South), October 27-28, 2019.IEEE, 2019, pp. 2037–2040.

[31] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: representing scenes as neural radiance fields for view synthesis," Commun. ACM, vol. 65, no. 1, pp. 99–106, 2022.

[32] K. Schwarz, Y. Liao, M. Niemeyer, and A. Geiger, "GRAF: generative radiance fields for 3d-aware image synthesis," in Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual.2020.

[33] M. Niemeyer and A. Geiger, "GIRAFFE: representing scenes as compositional generative neural feature fields," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021. Computer Vision Foundation / IEEE, 2021, pp. 11453–11464.

[34] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating Videos with Scene Dynamics," in Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016.

[35] Barcelona, Spain, D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, Eds., 2016, pp. 613–621.

[36] Y. Zhou and T. L. Berg, "Learning temporal transformations from time-lapse videos," CoRR, vol. abs/1608.07724, 2016.

[37] W. Xiong, W. Luo, L. Ma, W. Liu, and J. Luo, "Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks," in 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018.Computer Vision Foundation / IEEE Computer Society, 2018, pp. 2364–2373.

[38] A. Clark, J. Donahue, and K. Simonyan, "Adversarial video generation on complex datasets," 2019.

[39] Y. Tian, J. Ren, M. Chai, K. Olszewski, X. Peng, D. N. Metaxas, and S. Tulyakov, "A good image generator is what you need for high-resolution video synthesis," 2021.

[40] I. Skorokhodov, S. Tulyakov, and M. Elhoseiny, "Stylegan-v: A continuous video generator with the price, image quality and perks of stylegan2," in IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022. IEEE, 2022, pp. 3616–3626.